

Description

CODING APPARATUS AND METHOD THEREOF FOR DETECTING AUDIO SIGNAL TRANSIENT

BACKGROUND OF INVENTION

[0001] 1. Field of the Invention

[0002] The present invention relates to a coding apparatus, and more specifically, to a coding apparatus capable of detecting transients of audio signals. The coding apparatus of the present invention can also determine a window block length while adopting frequency domain coding technology.

[0003] 2. Description of the Prior Art

[0004] At present, many coding apparatuses are based on different coding algorithms, such as MP3, AAC, WMA, and Dolby DigitalTM. These coding algorithms take into account the characteristics of the human auditory system, and have the advantage of high compression ratio

(generally more than ten times). These coding apparatuses adopt perceptual coding, frequency domain coding, window switching, dynamic bit allocation technologies, etc to eliminate unnecessary content of the original audio data.

[0005] Perceptual coding eliminates audio data unperceivable by the human auditory system for reducing the size of the original audio data. Generally speaking, a human being can only hear audio signals having a frequency ranging from 20Hz to 20KHz, and therefore any audio signals out of this range are not perceivable. In addition, if the audio data have a signal eminent in volume or in tone, a human listener is not able to perceive other signals close to that sound. This phenomenon is referred to as auditory mask. Thus, it is unnecessary to code those unperceivable signals while coding the audio data.

[0006] Frequency domain coding transforms time domain data with high relativity into nearly irrelative frequency domains in order to eliminate unnecessary content of audio data. The frequency domain coding generally includes transform coding and subband coding. Transform coding has higher resolution while subband coding has lower resolution but higher efficiency. Therefore it is possible to

combine these two kinds of coding methods to form a combined filter having different resolutions at different frequencies. However, the pre-echo effect is a problem in frequency domain coding. For instance, if the audio data contains sounds of rapidly increasing energy, quantization noise would increase. This results in the pre-echo effect. Both transform coding and subband coding suffer from the pre-echo effect, which occurs when the audio data is transformed back into the time domain.

[0007] A method, referred to as window switching, for eliminating the pre-echo effect is used to limit the error within a shorter period of time, so that the pre-echo effect are kept in the masking area. According to the window switching method, audio signals that are more stable are encoded with long windows, while signals including transients are encoded with short windows. However, the disadvantage of window switching is that more bits are required for storing audio data since the data needed to be encoded increases.

[0008] The quality of coding is greatly related to the allocation of bits in each subband. In order to allocate bits, it is necessary to analyze input signals continuously to allocate more bits into the subbands most perceivable by human

beings, and allocate fewer bits into the subbands less perceivable. Since the signals change continuously, human beings have different reactions under different conditions. This is referred to as dynamic bit allocation technology. A good allocation relies on a precise psychoacoustic model.

[0009] Fig.1 illustrates a conventional MPEG audio layer-3 signal coding method. First, a pulse code modulation (PCM) input signal 10 is divided into thirty-two frequency subbands of equal width by a polyphase filter bank 12. The polyphase filter bank 12 simply analyzes the relationship between frequency and time, but the frequency subbands of equal width cannot precisely reflect the characteristic of the human auditory system. In addition, neighboring frequency subbands have more overlapped parts so a modified discrete cosine transform (MDCT) 14 for compensation is required for the output of the polyphase filter bank 12. The MDCT 14 further divides the subbands for better spectrum resolution, and eliminates some overlapped parts generated by the polyphase filter bank 12. The MDCT 14 includes two windows of different block lengths, which are respectively an eighteen-sample long window and a six-sample short window. Since continuous windows are 50% overlapped, the length of the longwindow is

actually thirty-six and the length of the short window is actually twelve. When the audio signals are stable, the long window has a higher frequency resolution and a better compression ratio, while the short window provides a better time resolution. Since the long window has a lower time resolution, if transients occur in the long window, the quantization noise will spread to the whole block. In this case, the signals with less energy will suffer from the quantization noise because of lower masking effect and therefore cause distortion, such as the pre-echo effect. To avoid the pre-echo effect, the conventional MPEG audio signal coding adopts a psychoacoustic model 16 to detect the transients of the audio signals, and then performs the MDCT 14 with short windows. After transforming the input signal 10 to the frequency domains by using frequency domain coding technology, a quantization process 18 is performed according to the psychoacoustic model 16. Then a packing process 20 is performed to pack the audio data and output a bit stream output signal 22.

[0010] The window switching technology is a typical way to avoid the pre-echo effect when performing frequency domain coding, and thus a mechanism of detecting transients of the audio signals is important. Conventional MPEG audio

signal coding adopts the psychoacoustic model 16 to detect transients in the audio signals. Although the psychoacoustic model 16 is accurate, it is very complicated and has a higher cost as well. It is therefore not economical to adopt the psychoacoustic model 16 to detect transients of the audio signals in window switching.

SUMMARY OF INVENTION

[0011] It is therefore one of the objectives of the claimed invention to provide a coding apparatus capable of detecting transients of audio signals. In addition, the claimed invention provides a coding apparatus and method thereof capable of determining window block length in frequency domain coding to solve the above-mentioned problems.

[0012] According to the claimed invention, a coding apparatus for coding an input signal to an output signal is provided. The coding apparatus includes a polyphase filter bank, a transient detector connected to the polyphase filter bank, and a coding processing unit connected to the polyphase filter bank and the transient detector. The polyphase filter bank is for producing a plurality of subband samples according to the input signal, wherein different subband samples correspond to the input signal in different time intervals, and each subband sample includes a plurality of

frequency subbands. The transient detector is for determining a block length of a window including a plurality of weighted values. The transient detector includes a subband selector for selecting the plurality of subband samples as reference sample data, an energy calculator connected to the subband selector for calculating an energy sum of the frequency subbands of the reference sample data, a partition device connected to the subband selector and the energy calculator for dividing the reference sample data into several subsample data, each subsample data having at least a subband sample, and a comparator connected to the energy calculator for comparing an output value of the energy calculator with a first threshold value and outputting a signal representing the block length of the window according to the comparing result. The coding processing unit is for multiplying the plurality of frequency subbands by the plurality of weighted values of the window to produce a weighted result, and generating the output signal by a predetermined algorithm according to the weighted result.

[0013] The claimed invention further provides a method for coding an input signal to an output signal. The method includes: performing a subband coding process to produce

a plurality of subband samples according to the input signal, different subband samples corresponding to the input signal in different time intervals, each subband sample having a plurality of frequency subbands; performing a selection process to provide a window of a predetermined block length, the window including a plurality of weighted values, the selection process including selecting a plurality of subband samples from the plurality of subband samples as reference sample data, and determining a block length of the window according to an energy of the frequency subbands of the reference sample data in a predetermined frequency range; and performing a transform process to multiply the plurality of frequency subbands by the plurality of weighted values of the window determined in the selection process for producing a weighted result, and to produce the output signal by a predetermined algorithm according to the weighted result.

[0014] These and other objects of the present invention will be apparent to those of ordinary skill in the art after having read the following detailed description of the preferred embodiment that is illustrated in the various figures and drawings.

BRIEF DESCRIPTION OF DRAWINGS

- [0015] Fig.1 is a schematic diagram illustrating a conventional MPEG audio layer-3 signal coding method.
- [0016] Fig.2 is a schematic diagram of a coding apparatus according to an embodiment of the present invention.
- [0017] Fig.3 is a schematic diagram illustrating the subband samples.
- [0018] Fig.4 is a flowchart showing how the coding apparatus detects a transient according to another embodiment of the present invention.

DETAILED DESCRIPTION

- [0019] Fig.2 illustrates a schematic diagram of a coding apparatus 30 according to an embodiment of the present invention. The coding apparatus 30 is for coding a pulse code modulation (PCM) input signal 10 to a bit stream output signal 22. The coding apparatus 30 includes a polyphase filter bank 12, a transient detector 32, and a coding processing unit 34. The polyphase filter bank 12 produces a plurality of subband samples according to the input signal 10. Different subband samples correspond to the input signal 10 in different time intervals, and each subband sample includes a plurality of frequency subbands. The

coding processing unit 34 performs a modified discrete cosine transform (MDCT) to the plurality of frequency samples. The transient detector 32, which is connected to the polyphase filter bank 12 and the coding processing unit 34, can decide the block length of the window when the coding processing unit 34 performs the MDCT. The transient detector 32 includes a subband selector 36, an energy calculator 38, a partition device 40, and a comparator 42. The subband selector 36 selects a portion of the plurality of subband samples in a predetermined frequency range as a reference sample data. Then the energy calculator 38 calculates the energy sum of the reference sample data. Following that, the comparator 42 compares the energy sum of the reference sample data with a first threshold value. If the energy sum of the reference sample data is larger than the first threshold value, there is probably a transient existing in the reference sample data. In such case, the partition device 40 divides the reference sample data into several subsample data of equal width, each subsample data including at least a subband sample. Meanwhile, the energy calculator 38 calculates the energy difference of the frequency subband between two adjacent subsample data in a predetermined frequency range, and

transfers the energy difference value to the comparator 42 to compare with a second threshold value. If the energy difference value is larger than the second threshold value, then the coding processing unit 34 perform the MDCT with short windows, otherwise it will repeat until the partition device 42 finishes all possible subsample data combinations. If the energy difference between two adjacent subsample data is still less than the second threshold value, then the coding processing unit 34 performs the MDCT with long windows.

[0020] Fig.3 illustrates a schematic diagram of the subband samples according to this embodiment. The polyphase filter bank 12 outputs eighteen subband samples during a time period "t". Each subband sample includes thirty-two frequency subbands. The coding processing unit 34 performs the MDCT to each frequency subband in the overlapped section, i.e. thirty-six subband samples. The transient detector 32 detects where the transient occurs and the coding processing unit 34 performs the MDCT with either long windows or short windows. The predetermined frequency range normally means frequency between a start subband and a coding limit subband. The subband selector 36 selects a frequency subband in this frequency

range as reference sample data 50. The start subband can be decided by experience or according to experimental results, and can be, for example, the first subband or a high frequency subband. In this embodiment, the frequency of the start subband is about 4kHz. On the other hand, the frequency of the coding limit subband has to be decided by coding criteria. Since the bit rate and the bandwidth are limited, the coding apparatus may discard some information of high frequency subbands. If no information is discarded, the last subband is the coding limit subband.

[0021] After the reference sample data 50 is selected, the energy calculator 38 calculates the energy sum contained in the reference sample data 50, and the comparator 42 decides whether or not to detect the reference sample data 50. The partition device 40 divides the reference sample data 50 into several equal width subsample data. Then the energy calculator 38 calculates the energy difference between two adjacent subsample data, and the comparator 42 decides the block length of the window. For example, the energy calculator 38 calculates the energy sum of the reference sample data 50 selected by the subband selector 36. If the energy sum is larger than -60dB , a transient

may exist in the reference sample data 50. In this case, the partition device 40 then divides the subband samples of the reference sample data 50 into six groups of subsample data of equal width. Then the energy calculator 38 calculates the energy difference between two adjacent groups of subsample data, and transfers the result to the comparator 42. If the energy difference between two adjacent subsample data is not larger than 20dB, then no transient actually occurs between the two adjacent subsample data. In such case, the partition device 40 re-divides the subband samples of the reference sample data 50 into three groups of equal width subsample data. Then the energy calculator 38 calculates the energy difference of the subsample data between two adjacent groups of subsample data, and the comparator 42 determines whether the energy difference is larger than 12dB. If the energy difference is larger than 12dB, then it is determined that there is a transient and short windows are selected. If the energy difference is not larger than 12dB, then long windows are selected.

[0022] Fig.4 is a flowchart illustrating how the coding apparatus 30 detects the transient in another embodiment of the present invention. Primarily, a subband coding process is

performed to generate a plurality of subband samples corresponding to the input signal 10. Different subband samples correspond to the input signal 10 in different time intervals, and each subband sample includes a plurality of frequency subbands. Then a selection process is performed for deciding the window block length for the next process. The window includes a plurality of weighted values. In the selection process, a plurality of subband samples are selected from the plurality of subband samples as reference sample data, and the window block length is decided according to the energy sum of the frequency subbands of the reference sample data in the predetermined frequency range. Finally a transform process is performed to multiply the plurality of frequency subbands by the plurality of weighted values decided in the selection process for generating a weighted result, and output the output signal by the MDCT according to the weighted result.

[0023] Detailed steps of detecting the transient according to this embodiment are illustrated as follows:

[0024] Step 110: Start.

[0025] Step 120: Is the energy sum of the reference sample data larger than a first threshold value? If yes, proceed step

130, otherwise, proceed step 170.

[0026] Step 130: Divide the reference sample data into several equal width subsample groups and calculate the energy of each subsample group.

[0027] Step 140: Is the energy difference between two adjacent subsample groups larger than a second threshold value? If yes, proceed step 160, otherwise, proceed step 150.

[0028] Step 150: Can the reference sample data be divided into different subsample data? If yes, return to step 130, otherwise, proceed step 170.

[0029] Step 160: Transform with short windows, then proceed step 180.

[0030] Step 170: Transform with long windows, then proceed step 180.

[0031] Step 180: End.

[0032] Please note that if the energy difference between adjacent subsample groups is not larger than the second threshold value in step 140 and the reference sample data can be divided into different subsample data, the reference sample data will be divided into several different subsample groups in step 130, and compared with the second threshold value again. However, since the subsample groups are different, the second threshold value may be

changed during the iterative steps of detecting the transient.

[0033] In comparison with the prior art, the present invention provides a coding apparatus and method thereof for deciding the window block length when performing the MDCT. It is worth noting that the present invention determines whether a transient exists by comparing the energy of the frequency subbands generated in encoding. Therefore, the present invention is more economical than the prior art, which uses the psychoacoustic model.

[0034] Those skilled in the art will readily observe that numerous modifications and alterations of the device may be made while retaining the teachings of the invention. Accordingly, the above disclosure should be construed as limited only by the metes and bounds of the appended claims.